

Notice of Allowability

Application No.

10/605,631

Examiner

USMAAN SAEED

Applicant(s)

FRANCIOSA ET AL.

Art Unit

2166

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☒ This communication is responsive to the amendment dated 06/18/2008.
2. ☒ The allowed claim(s) is/are 1-4, 7-14 and 17-20 (renumbered as 1-16).
3. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some* c) ☐ None of the:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. ____.
3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).
- * Certified copies not received: ____.

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.

THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

4. ☐ A SUBSTITUTE OATH OR DECLARATION must be submitted. Note the attached EXAMINER'S AMENDMENT or NOTICE OF INFORMAL PATENT APPLICATION (PTO-152) which gives reason(s) why the oath or declaration is deficient.
5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
- (a) ☐ including changes required by the Notice of Draftsperson's Patent Drawing Review (PTO-948) attached
- 1) ☐ hereto or 2) ☐ to Paper No./Mail Date ____.
- (b) ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date ____.
- Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).
6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☒ Notice of References Cited (PTO-892)
2. ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3. ☐ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date ____
4. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material
5. ☐ Notice of Informal Patent Application
6. ☒ Interview Summary (PTO-413), Paper No./Mail Date 8/25/08.
7. ☒ Examiner's Amendment/Comment
8. ☒ Examiner's Statement of Reasons for Allowance
9. ☐ Other ____.

DETAILED ACTION

1. This communication is in response to the amendment filed on 6/18/2008.

After thorough search and examination of the present application and in light of the prior art made of record, claims 1-4, 7-14 and 17-20 (renumbered as 1-16) are allowed.

Claims 5-6 and 15-16 have been cancelled.

EXAMINER'S AMENDMENT

2. An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it **MUST** be submitted no later than the payment of the issue fee.

Authorization for this examiner's amendment was given in a telephone interview with Attorney, Kevin M. Dunn, Registration No. 52,842 on August 25, 2008.

Please amend the claims, which were filed on 6/18/2008 with new versions as follows:

1. (Previously Presented) A method for computing a measure of similarity between a first (or input) document and one or more disparate (or search results) documents, comprising:

(a) receiving a first document and identifying the best keywords in the text by recognizing rare and uncommon keywords, including keywords that belong to one or more domain specific or subject matter specific dictionary;

(b) identifying documents similar to the first document using a query by formulating wrappers using the list of the best keywords identified in the first document that also appear in a DS dictionary;

(c) receiving a first list of rated keywords extracted from the first document and a list of rated keywords extracted from each of the one or more disparate documents;

(d) comparing the first list of rated keywords to the list of rated keywords from each of the one or more disparate documents to determine whether the first document forms part of the one or more disparate documents using a first computed percentage indicating what percentage of keyword ratings in the first list also exist in the list of at least one of the one or more disparate documents;

(e) verifying inclusion of the first document in the one or more disparate documents by computing a second percentage for each of the one or more disparate documents indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the first list also exist in the list for at least one of the one or more disparate documents when the first computed percentage indicates that the first document is included in at least one of the one or more disparate documents;

(f) using the first computed percentage to specify the measure of similarity when the computed second percentage for at least one of the one or more disparate documents is greater than the first computed percentage;

(g) ranking the one or more disparate documents based on the percentage computed indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the first list also exist in the list for at least one of the one or more disparate documents;

(h) if the first computed percentage does not indicate that the first document is included in the second document, computing a third percentage using the Jaccard similarity distance measure, wherein if said Jaccard similarity distance measure is greater than about 90 percent, the second document is identified as a revision of the first document, and if the Jaccard similarity distance measure is less than about 90 percent, said measure is a similarity measure between said first and second document; and

(i) if the third computed percentage indicates that the first document is a revision of the second document, computing a fourth percentage indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the second list also exist in the first list.

2. (Original) The method according to claim 1, wherein the second percentage at (c) is computed by giving weight only to those keywords and their set of neighboring keywords in the first list that match in the second list and a threshold percentage of the keywords in their set of neighboring keywords.

3. (Original) The method according to claim 2, wherein the second percentage at (c) is

Art Unit: 2166

computed by giving full weight to those keywords in the first list of rated keywords that cannot be accurately identified as having a complete set of neighboring keywords in the second set of keywords.

4. (Original) The method according to claim 2, wherein the threshold percentage is reduced when the first list of rated keywords is identified using OCR.

5. (Cancelled).

6. (Cancelled).

7. (Previously Presented). The method according to claim 1, further comprising the fourth computed percentage to specify the measure of similarity except when: (i) the fourth computed percentage is greater than the second computed percentage; (ii) the first list of rated keywords is identified using OCR; (iii) the fourth computed percentage is greater than fifty percent; and (iv) less than twenty percent of the keywords in the first list of keywords are in the second list of keywords.

8. (Original) The method according to claim 1, wherein the first computed percentage indicates that the first document is included in the second document when the percentage defined by ratio of $\text{Sum1}/\text{Sum2}$ is greater than approximately ninety percent, where: D1 is the number of keywords in first list of keywords; D2 is the number of

Art Unit: 2166

keywords in the second list of keywords; Sum1 is the sum of the weights of keywords that appear in D1 that also appear in D2; Sum2 is the sum of the weights of keywords in D1.

9. (Original) The method according to claim 1, wherein the first list of rated keywords includes one or more keywords translated from a second language different from a first language that is identified as being a primary language of the first document.

10. (Original) The method according to claim 1, wherein the first document is a portion of the second document.

11. (Currently amended) A computer-based system for computing a measure of similarity between a first (or input) document and one or more (or search results) documents, comprising:

a processor;

(a) means for receiving a first document and identifying the best keywords in the text by recognizing rare and uncommon keywords, including keywords that belong to one or more domain specific or subject matter specific dictionary;

(b) means for identifying documents similar to the first document using a query by formulating wrappers using the list of the best keywords identified in the first document that also appear in a DS dictionary;

(c) means for receiving a first list of rated keywords extracted from the first document and a list of rated keywords extracted from each of the one or more disparate documents, wherein keywords are rated at least in part by a relevant weight from their associated document language;

(d) means for comparing the first list of rated keywords to the list of keywords from each of the one or more disparate documents to determine whether the first document forms part of the one or more disparate documents using a first computed percentage indicating what percentage of keyword ratings in the first list also exist in the list of at least one of the one or more disparate documents;

(e) means for verifying inclusion of the first document in the one or more disparate documents by computing a second percentage for each of the one or more disparate documents indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the first list also exist in the list for at least one of the one or more disparate documents when the first computed percentage indicates that the first document is included in at least one of the one or more disparate documents; and

(f) means for using the first computed percentage to specify the measure of similarity when the computed percentage for at least one of the one or more disparate documents is greater than the first computed percentage;

(g) means for ranking the one or more disparate documents based on the percentage computed indicating what percentage of keyword ratings along with a set of

Art Unit: 2166

their neighboring keyword ratings in the first list also exist in the list for at least one of the one or more disparate documents;

(h) if the first computed percentage does not indicate that the first document is included in the second document, means computes a third percentage using the Jaccard distance measure, wherein if said Jaccard similarity distance measure is greater than about 90 percent, the second document is identified as a revision of the first document, and if the Jaccard similarity distance measure is less than about 90 percent, said measure is a similarity measure between said first and second document; and

(i) if the third computed percentage indicates that the first document is a revision of the second document, means computes a fourth percentage indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the second list also exist in the first list.

12. (Original) The system according to claim 11, wherein the second percentage at (c) is computed by said computing means by giving weight only to those keywords and their set of neighboring keywords in the first list that match in the second list and a threshold percentage of the keywords in their set of neighboring keywords.

13. (Original) The system according to claim 12, wherein the second percentage at (c) is computed by said computing means by giving full weight to those keywords in the first

Art Unit: 2166

list of rated keywords that cannot be accurately identified as having a complete set of neighboring keywords in the second set of keywords.

14. (Original) The system according to claim 12, wherein the threshold percentage is reduced when the first list of rated keywords is identified using OCR.

15. (Cancelled)

16. (Cancelled)

17. (Original) The system according to claim 16, further comprising means for using the fourth computed percentage to specify the measure of similarity except when: (i) the fourth computed percentage is greater than the second computed percentage; (ii) the first list of rated keywords is identified using OCR; (iii) the fourth computed percentage is greater than fifty percent; and (iv) less than twenty percent of the keywords in the first list of keywords are in the second list of keywords.

18. (Original) The system according to claim 11, wherein the first computed percentage indicates that the first document is included in the second document when the percentage defined by ratio of $\text{Sum1}/\text{Sum2}$ is greater than approximately ninety percent, where: D1 is the number of keywords in first list of keywords; D2 is the number of keywords in the second list of keywords; Sum1 is the sum of the weights of keywords

Art Unit: 2166

that appear in D1 that also appear in D2; Sum2 is the sum of the weights of keywords in D1.

19. (Original) The system according to claim 11, wherein the first list of rated keywords includes one or more keywords translated from a second language different from a first language that is identified as being a primary language of the first document.

20. (Currently Amended) An article of manufacture for computing a measure of similarity between a first (or input) document and one or more disparate (or search results) documents, the article of manufacture comprising computer usable storage media including computer readable instructions embedded therein that causes a computer to perform a method, wherein the method comprises:

- (a) receiving a first document and identifying the best keywords in the text by recognizing rare and uncommon keywords, including keywords that belong to one or more domain specific or subject matter specific dictionary;

- (b) identifying documents similar to the first document using a query by formulating wrappers using the list of the best keywords identified in the first document that also appear in a DS dictionary;

- (c) receiving a first list of rated keywords extracted from the first document and a second list of rated keywords extracted from the second document;

(d) using the first and second lists of rated keywords to determine whether the first document forms part of the second document using a first computed percentage indicating what percentage of keyword ratings in the first list also exist in the second list;

(e) verifying inclusion of the first document in the second document computing a second percentage indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the first list also exist in the second list when the first computed percentage indicates that the first document is included in the second document;

(f) using the first computed percentage to specify the measure of similarity when the second computed percentage is greater than the first computed percentage;

(g) if the first computed percentage does not indicate that the first document is included in the second document, computing a third percentage using the Jaccard distance measure, wherein if said Jaccard similarity distance measure is greater than about 90 percent, the second document is identified as a revision of the first document, and if the Jaccard similarity distance measure is less than about 90 percent, said measure is a similarity measure between said first and second document; and

(h) if the third computed percentage indicates that the first document is a revision of the second document, computing a fourth percentage indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the second list also exist in the first list, the fourth computed percentage is used to specify the measure of similarity except when: (i) the fourth computed percentage is greater than the second computed percentage; (ii) the first list of rated keywords is identified using OCR; (iii) the

Art Unit: 2166

fourth computed percentage is greater than fifty percent; and (iv) less than twenty percent of the keywords in the first list of keywords are in the second list of keywords.

Reason for Allowance

3. The prior art made of record does not teach or fairly suggest the combination of elements, as recited in independent claims 1, 11 and 20.

More specifically, the prior art of records does not specifically suggest the combination of “if the first computed percentage does not indicate that the first document is included in the second document, computing a third percentage using the Jaccard distance measure, wherein if said Jaccard similarity distance measure is greater than about 90 percent, the second document is identified as a revision of the first document, and if the Jaccard similarity distance measure is less than about 90 percent, said measure is a similarity measure between said first and second document; and if the third computed percentage indicates that the first document is a revision of the second document, computing a fourth percentage indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the second list also exist in the first list” in combination with all the other limitations in the independent claims 1, 11, and 20.

These features together with other limitations of the independent claim are novel and non-obvious over the prior art of record. The dependent claims 2-4, 7-10, 12-14, and 17-19 being definite, enabled by the specification, and further limiting to the independent claims, are also allowable.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

Contact Information

4. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Usmaan Saeed whose telephone number is (571)272-4046. The examiner can normally be reached on M-F 8-5.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Hosain Alam can be reached on (571)272-3978. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Usmaan Saeed
Patent Examiner
Art Unit: 2166

Application/Control Number: 10/605,631
Art Unit: 2166

Page 14

Hosain Alam
Supervisory Patent Examiner

US
August 26, 2008

/Hosain T Alam/

Supervisory Patent Examiner, Art Unit 2166